International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 12, pp 20-23, December-2024

https://doi.org/10.47001/IRJIET/2024.812003

Exposing to Infiltrate SCADA Using Whale Algorithm and Support Carrier

[1]Jenan Jader Msad, [2]Dhulfikar Dhurgham Husam, [3]Mustafa Husham Abbas

[1,2,3]Department of Computer Science, University of Al-Furat Al-Awsat, Najaf, Iraq

Authors E-mail: [1]jenan.jader@atu.edu.iq, [2]talfekar@atu.edu.iq, [3]mustafa.abbas@atu.edu.iq

Abstract **- In today's growing cyber attack, where the critical infrastructure of a country's communications and services can quickly be eliminated by hostile attacks, the protection of vital infrastructure and advanced cyber security is needed at all times. In fact, security defects for such systems can lead to widespread destruction and consequences in different layers of society. The SCADA system (information control and collection control system (process control system is computer -based. It monitors and interacts with the physical process of the remote road. This system collects information about the physical process conditions; the system is widely used in large industries such as petrochemicals, water distribution and nuclear power plants. Monitoring the proper performance of the process is provided on large control panels that allow the operator to allow the operator. With the introduction of telemetry, the conditions for connecting the equipment used in this infrastructure are created. In fact, the term SCADA refers to a technology that creates conditions for controlling and monitoring such infrastructures from a central control room. For example, this technology can be used in water distribution facilities to check the water level in the water storage tank, as well as monitor the flow and pressure in the pipe. The SCADA system enables the intelligent collection of information from various parts for the user and provides real-time monitoring of all processes, and may also warn in the event of an error and the operator can react by sending commands to work equipment compared to process changes.**

Keywords: SCADA system, whale algorithm, machine learning, classification, water distribution facilities.

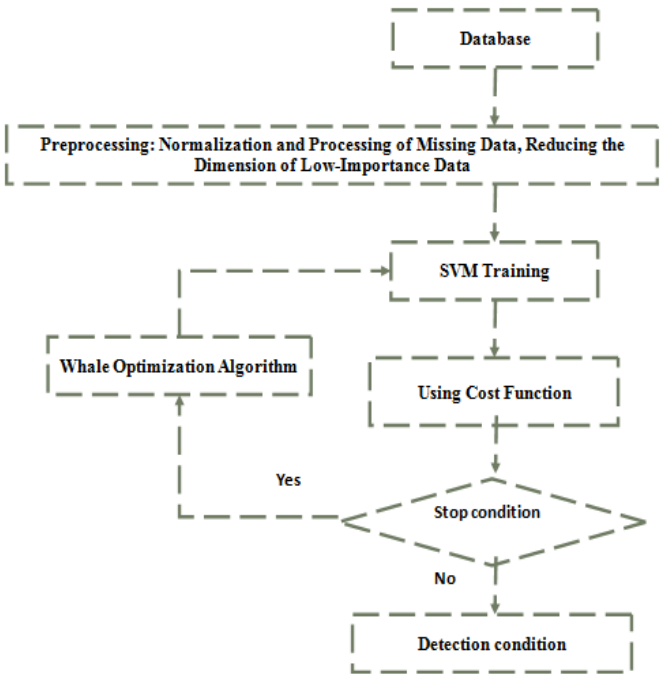
I. Introduction

The industrial control system is distributed in the form of cyber physical systems and is considered as an important component of many infrastructures in each country, with penetration and failure in these systems leading to irreparable damage. One of the most effective methods of intrusion detection in industrial control systems is the use of machine learning -based methods because they can detect new data based on the concepts taught from previous data. In this thesis, we identify the influence in the industrial control system. The purpose and motivation of this study is to increase the accuracy of diagnosis and reduce false positive alert in the field of intrusion detection. For this purpose, the back -up machine was used to achieve the purpose of this research with the whale's algorithm.

II. Necessity of Research

Control system is distributed in the form of cyber physical systems and is considered as an important component of many infrastructures in each country, with penetration and failure in these systems leading to irreparable damage. One of the most effective methods of intrusion detection in industrial control systems is the use of machine learning -based methods because they can detect new data based on the concepts taught from previous data. In this thesis, we identify the influence in the industrial control system. The purpose and motivation of this study is to increase the accuracy of diagnosis and reduce false positive alert in the field of intrusion detection. For this purpose, the back -up machine was used to achieve the purpose of this research with the whale's algorithm.

III. Proposed Method



International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 12, pp 20-23, December-2024

https://doi.org/10.47001/IRJIET/2024.812003

IV. Description of Fundamental Component Analysis

In this work, in order to increase classification accuracy and also reduce false positive warning for detection of intrusion in industrial control systems, we intend to use a support vector machine optimized with whale algorithm as an innovation aspect of this work. In fact, we'll use the whale algorithm to adjust the value of the two C parameters and the kernel value described below:

The analysis of the basic components of a conversion in a vector space, which is mostly used to reduce the dimensions of the data set.

The analysis of basic components can be used to reduce data dimensions, thus maintaining parameters of the data set that have the maximum effect on variance.

The method of analyzing the basic components has two functions: its first function is to identify the factors directly and without estimating subscriptions from the correlation matrix. Using this method to determine the maximum variation variable value, their linear composition is calculated.

The first parameter specifies the highest variation of variables. The second parameter then explains the highest amount of variance in the variables after the first parameter and continues to the end.

Another function of the analysis of the basic components is to convert a set of measured variables into a set of orthogonal linear compounds with maximum variance.

The steps of the analysis algorithm are the basic component as follows: Data normalization.

V. Backup Vector Machine Training and Optimization Using Whale Algorithm

In this work, we intend to use the backup vector classroom to classify data in two normal classes or penetrate.

The basis of this category is that it first transfers nonlinear data to the linear environment using the kernel function and then in linear environment seeks a separating plate that has the most distance from both classes.

The backup vector car classroom also uses a nonlinear kernel function to transfer data to a linear separate environment.

The backup vector car category has many kernel functions including: linear kernel, polynomial, Gaussian and more. One of the most important and important kernels in this algorithm is the Gaussian kernel,

$$K(x.y) = \exp\left(-\frac{||x-y||^2}{2\sigma^2}\right)$$

Training and optimization of the support vector machine using the Whale Classifier algorithm of the backup vector machine, after transferring the data to separable space by kernel functions, the algorithm is looking for a separator plane cloud that has the greatest distance from both available classes. Therefore, the objective function of the support vector machine is defined as the following equation.

$$Max\ \frac{1}{2}\ ||w||^2 + C\sum_{i=1}^{N}\epsilon_i$$

$$s.t.\, y_i(w^T\ \emptyset\ (x_i) + b) - 1 + \epsilon_i \geq 0 \forall i = 1,2,\ldots..,N$$

The backup vector machine has a set of components that play an important role in the performance of the classification, and if we do not have a proper understanding of these components, it will certainly not be appropriate for them and as a result the classification cannot work well. To achieve the best performance, we need to consider the right values for its components. The backup vector machine is a basic component for linear state (C) or the same cost and two basic parameters for nonlinear C) and parameter related to nonlinear kernels). Parameter C is a positive constant semester that can control the border.

VI. Analysis of the Results of the Proposed Method

In this part of the thesis, we will analyze and analyze the results of the proposed method to detect penetration in the industrial control system. As we said, after receiving the data from the data set, these data were normalized using a 4-1 relationship and all data in the dataset were processed using the average value of 3 neighboring. The result of this step is shown in Figure 1. As it is clear in this figure, all data are limited to 1 to 1, and there is no trace of the data available in the data set.

	1	2	3	4	5	6	7
1	0	2.6198e-09	5.2396e-09	2.6198e-08	1.2863e-06	0	0
2	0	5.2396e-09	7.8594e-09	2.6198e-08	3.8249e-07	0	0
3	0	2.6198e-09	1.0479e-08	1.5719e-08	0	0	0
4	0	2.6198e-09	1.3099e-08	2.6198e-08	6.0779e-07	2.1359e-05	0
5	0	2.6198e-09	1.3099e-08	2.6198e-08	5.2134e-07	1.1003e-06	0
6	0	2.6198e-09	1.0479e-08	5.2396e-09	0	0	0
7	0	2.6198e-09	1.0479e-08	1.5719e-08	0	0	0
8	0	2.6198e-09	1.0479e-08	1.5719e-08	0	0	0
9	0	2.6198e-09	1.5719e-08	1.5719e-08	0	0	0
10	0	2.6198e-09	1.0479e-08	1.5719e-08	0	0	0
11	0	2.6198e-09	1.0479e-08	5.2396e-09	0	0	0
12	0	2.6198e-09	1.0479e-08	1.5719e-08	0	0	0
13	0	2.6198e-09	1.3099e-08	2.6198e-08	7.5188e-07	5.8972e-06	0
14	0	2.6198e-09	5.2396e-09	2.6198e-08	8.7501e-07	0	0
15	0	2.6198e-09	1.8339e-08	1.5719e-08	0	0	0

Figure 1: The result of the data pre-processing

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 12, pp 20-23, December-2024

https://doi.org/10.47001/IRJIET/2024.812003

In the next step, as we said, we eliminated the dimensions of the data using the basic component analysis method and the dimensions of the data set converted from 40*5000 to 57*37 and 3 dimensions of each sample with the basic component analysis method we removed. After dividing the data with a ratio of 70 to 30, we taught and evaluate the backup vector car classroom as well as the whale's classification. The kernel used in the backup vector car class was of the Gaussian type, and the Gaussian kernel scale value and the margin adjustment parameter in the goal of the backup vector machine were adjusted with the whale algorithm. The following is the specifications of the whale algorithm parameters in Table 1.

Table 1: Specification of the Whale Optimizer Algorithm

Initial population number	10
Number of iterations	50 repetitions
Setting parameter	Based on Gaussian kernel scale and margin
Repetition Stop	Reach 50 replications
Proportional function	Condition Classification error value

The following is the process of optimizing the backup vector machine parameters for 50 repeats.

In this graph, the horizontal axis shows the number of repetitions and the vertical axis of the error. As it is clear in the figure, this diagram is a descending chart and states that the error in each repetition is declining, indicating that in each repetition the amount determined for the backup vector parameters is optimized.

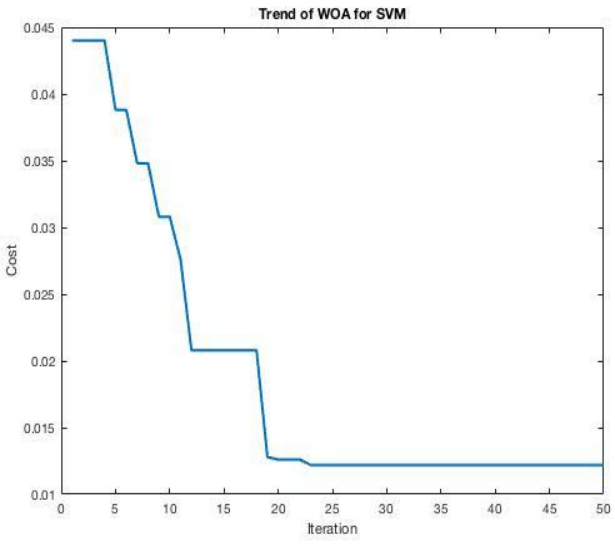


Figure 2: The process of optimizing the parameters of the support vector machine for 50 repetitions

We will examine the results for the training data in the form of characteristic curve graph and also the distorted matrix.

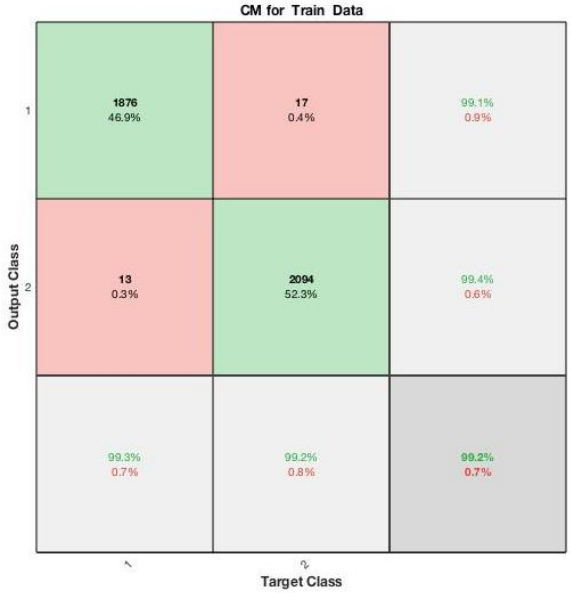


Figure 3: Classification data clutter matrix

In this matrix each input represents a criterion and an evaluation parameter. To analyze this matrix we will show in Figure 3 that each of the inputs of this matrix represents what the parameter is.

VII. Conclusion

The industrial control system is today in the form of cyber physical systems and as a component. The crucial and critical of many of the infrastructure of any country is considered, and ultimately infiltration and failure in these systems can lead to irreparable damage. Traditionally, penetration detection systems are analyzed by researchers (security analysts). They evaluate the warnings and decide what operation to do. However, this is a very difficult and time consuming job because the number of warnings can be very high and the environment may change rapidly. One of the most effective methods of intrusion detection in industrial control systems is the use of machine learning -based methods because they can detect new data based on the concepts taught from previous data.

In this thesis, we identify the influence in the industrial control system. The purpose and motivation of this study is to increase the accuracy of diagnosis and reduce false positive alert in the field of intrusion detection. For this purpose, the back -up machine was used to achieve the purpose of this research with the whale's algorithm. The vector car is one of the most powerful machine learning classrooms. In this thesis, using the whale algorithm to determine the optimal value of

ISSN (online): 2581-3048
Volume 8, Issue 12, pp 20-23, December-2024
https://doi.org/10.47001/IRJIET/2024.812003

the Gucci kernel scale parameters as well as the vector machine algorithm margin value parameter

Determining the appropriate value of the Gaussian kernel scale parameter makes the data properly transferred to the line able space, as well as determining the appropriate value of the border adjustment parameter enables the support vector machine to properly detect the optimized separator plate. Finally, we were able to reach 98% for the accuracy criterion, which was more accurate than the basic article. It should be noted that the dataset of the penetration system of industrial control systems used in this thesis relates to the gas pipeline system used in the basic article.

REFERENCES

[1] Nguyen, D. D., Le, M. T., & Cung, T. L. (2024). Improving intrusion detection in SCADA systems using stacking ensemble of tree-based models. *Bulletin of Electrical Engineering and Informatics*, 11(1), 119-127.

[2] Alimi, O. A., Ouahada, K., Abu-Mahfouz, A. M., Rimer, S., & Alimi, K. O. A. (2022, April). Supervised learning based intrusion detection for SCADA systems. *In 2024 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development (NIGERCON)* (pp. 1-5). IEEE.

[3] Alshadoodee, H. A. A., Mansoor, M. S., Kuba, H. K., & Gheni, H. M. (2024). The role of artificial intelligence in enhancing administrative decision support systems by depends on knowledge management. *Bulletin of Electrical Engineering and Informatics*, 11(5).

[4] Sheng, C., Yao, Y., Fu, Q., & Yang, W. (2021). A cyber-physical model for SCADA system and its intrusion detection. *Computer Networks*, 185, 107677.

[5] Altaha, M., Lee, J. M., Aslam, M., & Hong, S. (2021). An Autoencoder-Based Network Intrusion Detection System for the SCADA System. *J. Commun.*, 16(6), 210-216.

[6] Alimi, O. A., Ouahada, K., Abu-Mahfouz, A. M., Rimer, S., & Alimi, K. O. A. (2021). A review of research works on supervised learning algorithms for SCADA intrusion detection and classification. *Sustainability*, 13(17), 95-97.

[7] Rakas, S. V. B., Stojanović, M. D., & Marković-Petrović, J. D. (2020). A review of research work on network-based scada intrusion detection systems. *IEEE Access*, 8, 93083-93108.

[8] Khan, I. A., Pi, D., Khan, Z. U., Hussain, Y., & Nawaz, A. (2019). HML-IDS: A hybrid-multilevel anomaly prediction approach for intrusion detection in SCADA systems. *IEEE Access*, 7, 89507-89521.

[9] Abou El-Ela, A. A., El-Sehiemy, R. A., & El-Shebiny, A. M. (2019). Review of SCADA System for Distribution Power System Automation. *ERJ. Engineering Research Journal*, 42(2), 93-98.

[10] Sheng, C., Yao, Y., Fu, Q., & Yang, W. (2021). A cyber-physical model for SCADA system and its intrusion detection. *Computer Networks*, 185, 107677.

[11] Ujvarosi, A. (2016). Evolution of SCADA systems. Bulletin of the Transilvania University of Brasov. *Engineering Sciences. Series* I, 9(1), 63.

[12] Sazid, M. M., Nuhel, A. K., Shakib, M. A. A., Parvez, N., & Billah, M. (2022, April). Developing a Low-cost SCADA System for Industrial Application. *In 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 227-232). IEEE.

[13] Gao, J., Gan, L., Buschendorf, F., Zhang, L., Liu, H., Li, P., ... & Lu, T. (2020). Omni SCADA intrusion detection using deep learning algorithms. *IEEE Internet of Things Journal*, 8(2), 951-961.

[14] Upadhyay, D., Manero, J., Zaman, M., & Sampalli, S. (2021). Intrusion detection in SCADA based power grids: Recursive feature elimination model with majority vote ensemble algorithm. *IEEE Transactions on Network Science and Engineering*, 8(3), 2559-2574.

Citation of this Article:

Jenan Jader Msad, Dhulfikar Dhurgham Husam, & Mustafa Husham Abbas. (2024). Exposing to Infiltrate SCADA Using Whale Algorithm and Support Carrier. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 8(12), 20-23. Article DOI https://doi.org/10.47001/IRJIET/2024.812003

****

